Network Visualisation as a Citator User Interface

Neil du Toit*
Data Scientist, African Legal Information Institute, Democratic Governance and Rights Unit, University of Cape Town

Abstract. This paper describes the benefits of using network visualisation as a user interface to Citators. I show that three types of search queries that arise in legal research can be framed as querying the existence of certain composite citation relationships, and that these queries can be evaluated by the user significantly faster using network visualisation. I discuss further implications of network visualisation for exploratory data analysis, and I present our work on the AfricanLII Citator application, currently in beta, which provides such an interface to a citator that indexes case law from 15 countries in Africa.

Keywords: Network Visualisation, Jurisprudence, Stare Decisis

1. Introduction

Common law jurisdictions are characterised by the legal principle of *stare decisis*, or "let the decision stand" (Duxbury, 2008). The principle of *stare decisis* requires that courts, when deciding a case before them, apply the principles established in previous decisions by courts of similar or higher authority, if they are applicable to the facts before them. In addition, practices for considering decisions by inferior courts and foreign jurisprudence as being non-binding, but having "persuasive" value, have developed, either as soft law, or in some cases encoded into legal instruments. A consequence of *stare decisis* is that courts in common law jurisdictions will regularly cite and discuss previous judgments from the same jurisdiction or from similar jurisdictions.

In 1807 a lawyer by the name of Simon Greenleaf relied upon an English decision in litigation which, unbeknown to him, had been overruled. He thus undertook to compile a table of "overruled cases" to avoid such predicaments in future (Patti, 1993). Since then, it has become the practice of many publishers to make available, along with the text of judgments, tables indicating which decisions were cited by the judgment, and which later decisions subsequently cited a decision.

The initial focus of citators was, in the tradition of Greenleaf, to ascertain whether or not a given decision remained good law. Early innovations in citator *interfaces* reflected this emphasis. One of the most successful citators, the Shepard Citator, is noteworthy particularly for its innovative presentation of the index. Shepard printed his citator onto gummed paper, so that notes to overruling cases could be pasted right into the reports next to the cases themselves (Dabney, 2007). This ultimately increased the speed at which citations could be found.

Over time, however, there has been a gradual recognition that citators serve a function beyond mere validation of authority. This progression is clear from Dabney's commentary on KeyCite (Dabney, 2007):

> "[...] the introduction of KeyCite was a sharp reminder that citators had a real role as a tool for doing legal research, rather than being the last step in the process."[1]

Contemporary accounts of citators tend to focus both on aspects of validation of search results, as well as legal research more generally. For instance, Washington University Law School characterises a citator as follows (Washington University, 2005):

> "Citators allow you to determine if your case is **still good law**, and it acts as a **research tool** allowing you find other cases (and other secondary materials) which cited your case [own emphasis]".

Despite the evolution of the role of citators towards general research tools, there has been a slower progression of citator interfaces to support this new role, with many citator interfaces still only presenting indexes in a tabular form. The Encyclopedia of Library and Information Science (Librarianship Studies, 2019), still defines a citator to be an "ordered list of cited articles along with a list of citing articles" (though this definition relates to citators generally).
Tabular representations of citation data are restricted to showing only those cases that are directly related to a given decision. However, within a network setting, one may often be interested not only in direct relationships between nodes, but more generally in certain composite relationships between nodes. Citations between cases can in a straightforward manner be considered as the edges in a graph in which decisions are the nodes, thereby forming a network of cases. Therefore, in the case of a jurisprudence network, one could conceivably ask, for example, whether there are any decisions which cite a decision which itself cites a given case.

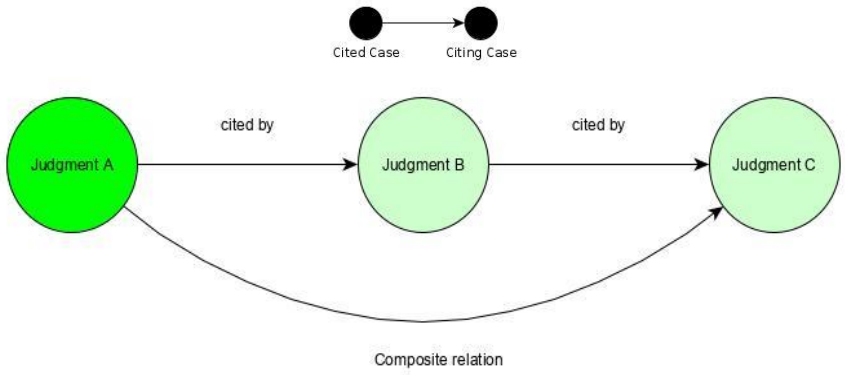

Figure 1. A composite citation relation

In this paper I will be using a number of smaller diagrams to illustrate

the network concepts discussed. I pause here to make a brief note regarding the directionality of the arrows. The convention adopted in this paper, which is also used in the Citator, is for citation arrows to represent the verb "is cited" rather than the verb "cites". Therefore, they point from the cited decision towards the citing decision. This passive form is perhaps less common in network visualization, but the approach does have two distinct advantages which convinced me to adopt it:

1. The directionality of the arrow can now be interpreted to represent a flow of information between judgments. This is particularly powerful in situations where the citing case has either approved or applied the cited case, since in these instances the directionality of the arrow can be interpreted as showing the direction in which legal precedent is moving between cases. One can therefore, for example, easily track how a new legal principle has moved from the decision which introduced it to later decisions.
2. This convention means that the earlier decision in time will always be at the start of the arrow and later decision in time will be at the end of the arrow. Since it is usually the case that more recent authority is preferred to earlier decisions, this convention allows the reader (or user in the case of the Citator), to simply follow a path along the arrows to find more recent decisions.

The arrow should thus not be seen as an action by one case on another, but rather as a flow of information from one case to another. This also provides an easy way to remember the convention: the case at the start of the arrow is always earlier in time.

In order to determine the existence of *composite* relationships between judgments, users working with tabular representations of citation networks have needed to locate and consult multiple tables in a time-consuming process.

The question therefore naturally arises as to whether such composite relationships between judgments are of any legal significance. I argue that they are, and specifically that they can greatly support users who are using a citator for research, not merely validation.

I present three examples of classes of common queries of importance to legal research, which can be viewed as queries on the existence of certain composite citation relationships. In particular, I discuss the legal value of searching for "second opinions" on judgments, for searching for recommended judgments, and for searching for related judgments. I show how these three classes of user queries can be best described as searching for certain composite citation relationships. Finally, I will discuss the process of "initial research" and its relationship with exploratory data analysis. I describe how visual interfaces are able to elucidate the nature of the jurisprudential landscape at a higher level of analysis. I then show how, and why, the time-consuming process of determining the existence of these composite relationships can be appreciably sped up by the use of network visualisation.

2. Composition of Relations

The meaning of a "composition of relations" can be illustrated by example (a precise definition appears in the appendix). Suppose that Peter is the father of Tom, and Tom is the father of Adam. Then I can infer a relationship between Peter and Adam: Peter is the father of the father of Adam. Furthermore, the composite relation has a meaningful significance within the domain of discourse. Indeed, this composite relation even has its own name: the grandfather relation.

Network representations of relations are most appropriate when the composition of those relations is able to result in composite relations which are themselves meaningful within the domain of discourse. Some composition relations are less obviously significant within the domain of discourse than the grandfather relation. Consider, for example, the Facebook friends network graph. Suppose that Peter is friends with Tom, and Tom is friends with Adam. Then we can infer a relationship between Peter and Adam. However, this connection more tenuous, and we usually simply refer to it as the "friend of a friend" relation.

Composition can be performed on relations that are not the same. For example, the composition of the "father" relation and the "brother" relation will yield the "uncle" relation. In general, however, composing different relations will not always lead to meaningful composite relations.

3. Composite Relations in Jurisprudence Research

A jurisprudence citation network consists of judgments, which are the nodes, and citations between judgments, which are the edges. To say that there is a relationship between two judgments means that one judgment cited the other.[2]

Citations can be further broken up and are usually categorised as follows: *Followed, approved, applied, considered, explained, distinguished, doubted, not applied* and *overruled* (ICLR, 2019).

I now provide three examples of user queries that can be represented as queries for composite relations of citations.

3.1 Recommender Systems

A popular feature of many search engines is recommender systems. Initially developed aggressively by Amazon, recommender systems are now prominent features on many e-commerce platforms and media streaming services.

Recommender systems work on the assumption that two users who have both "endorsed" a given search result will be likely to have

[2] This relationship is not symmetric. Furthermore, the citing case always occur later in time than the cited case. The jurisprudence network is therefore, more specifically, a directed acyclic graph.

similar preferences in general. For example, when Amazon notifies you that people who purchased the items in your basket also purchased another item, they are assuming that you are likely to have similar preferences to those people, and therefore may also be interested in the other product.

Recommender systems therefore require a system of tracking "endorsements" of some form. For Amazon, when you purchase a product you are "endorsing" it. When you listen to a song on Spotify, this is a form of endorsement. Endorsement doesn't necessarily mean that you approve of the item itself. It only implies that you considered it a useful search result for the search engine to return.

Endorsements can be directly applied to search results by ranking results higher when they have more endorsements. However, endorsements can also be used for recommender systems in those cases where it is reasonable to assume that users who endorse the same item will have similar preferences in general.

Citations between judgments can be considered to be an endorsement by the judge of the decision. This is the case even when a case is disapproved of or not followed, since the judge still found it necessary to comment on the decision, and therefore it remains a good search result when querying the given topic. The judge is of course not a user. But once can certainly still make the assumption that users will have similar preferences for search results to judges. Consider the situation in Figure 2:

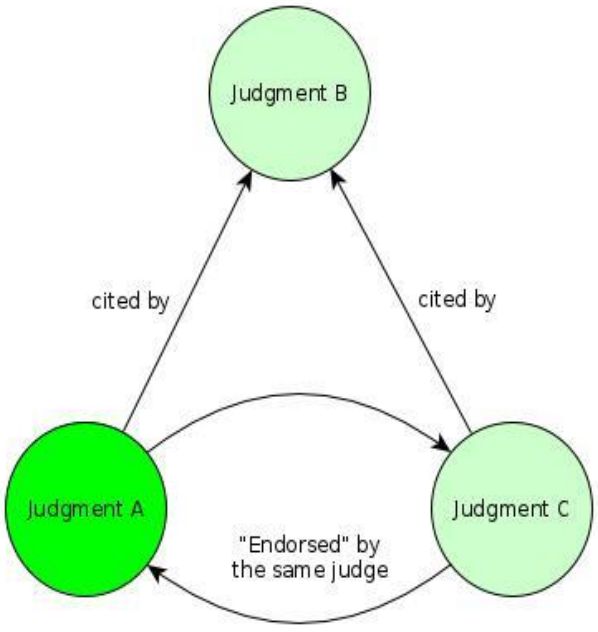

Figure 2. Citations as a recommender system

Suppose that a user is considering a judgment, "Judgment A". They notice that judgment A has been cited in judgment B. If judgment B also cited a different judgment C, then one could say to the user who is considering judgment A that "judges who cited this decision, also

cited judgment C". Or, more to the point: "you may also be interested in judgment C". A visual representation of a citator index allows the user to implicitly find "recommended" material in this manner, simply by looking for judgments that point to the same place as the one that they are currently reading.

This type of recommender system can be considered to be a search query on whether there exists a composite relation of "was cited by a decision that also cited". As such, the problem of finding such recommendations is an example of the problem of determining the existence of a certain composite relation.

3.2 Second Opinions

Consider the situation in Figure 3:

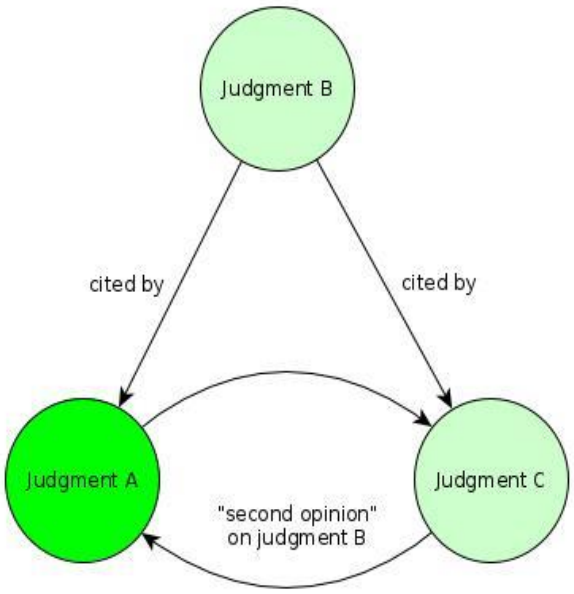

Figure 3. Obtaining a second opinion

Suppose that a user is reading a judgment, "Judgment A", which discusses another judgment, "Judgment B", at some length. The user might wish to know whether there are any *other* judgments which have expressed an opinion on "B". In this case the user can simply look for a decision which is *pointed to* by the same case that points to the case they are currently reading.

Looking for "second opinions" on a judgment can be described as a query on whether there exists a composite relation "cited a case which has also been cited by". As such the problem of obtaining a second opinion is an example of the problem of determining the existence of a particular composite relation.

3.3 Topic Clustering

A perusal of the different types of citation reveals that citations between judgements may signify that the judgments consider similar

topics. For example, when a court signals that it is following a previous decision, it is "expressing itself as bound by a previous decision of a court of coordinate or superior jurisdiction in a case where *the material facts were the same or substantially the same* [own emphasis]" (ICLR, 2019). When a court indicates that it is applying a previous decision, the facts of the case are different, but the court is expressing an opinion that the same legal principle is applicable to both.

The idea of finding judgments on similar topics using citations is not new to legal research. The strategy is sometimes described as "using one good case" to explore authority on a topic. The idea of citations as a basis for inferring a relation of ideas is also common to all types of citations, not merely jurisprudential citations. According to Encyclopedia of Legal Information Services, "the fact is that whenever a recent paper cites a previous paper there always exists a relation of ideas, between the two papers" (Librarianship Studies, 2019).

"Similarity" is a weakly transitive relation. If A is similar to B, and B is similar to C, then A and C are probably similar, but may not be. In practice, the link is usually strong enough to be useful up to 3 or 4 degrees of separation.

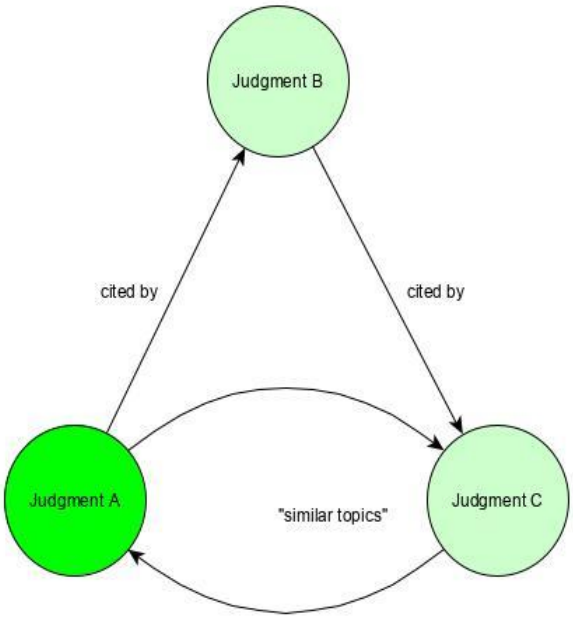

Figure 4. Topic clustering with citations

Users can therefore use citation networks in order to search for decisions with that deal with similar topics. When performing exploratory analysis, users may want to be shown a larger group of related decisions. In these instances, composition of citations is relevant, as users will need to look for more distantly related cases.

4. User Query Effort Comparison Between Tabular Representations and Visual Graph Representations of Networks.

In this section I will illustrate the process that must be followed to perform a graph traversal using tabular interfaces with an example, and compare this against a network visualisation interface.

Consider a legal researcher who is trying to determine whether or not there exists a line of decisions between two judgments, let's call them *S v Alice* and *S v Grace.* At her disposal she has six tables which have, in their header, the name of a decision and, in their body, a list of decisions which have cited that decision.

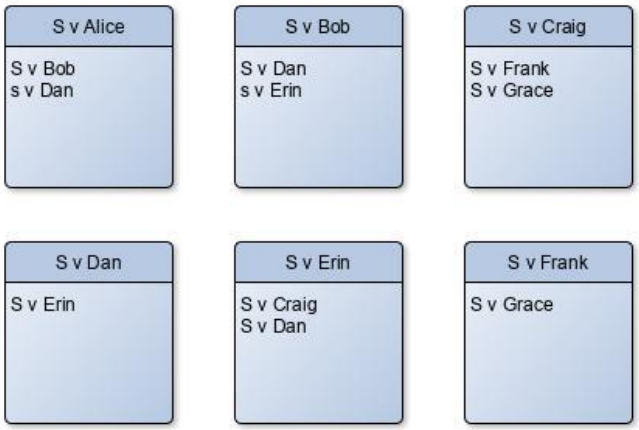

Figure 5. Tables of citing cases

Study the tables and try to deduce, based on these tables, whether there is a line of decisions between *S v Alice* and *S v Grace*. I should pause here to note that rarely would a researcher be so fortunate to have all relevant cases on hand, as well as only having to consider at most two citing cases. Notwithstanding the generosity of the example, I'm nonetheless certain that the reader found the determination to be at least somewhat unwieldy.

Now let's consider the same data, except presented in a visual network representation:

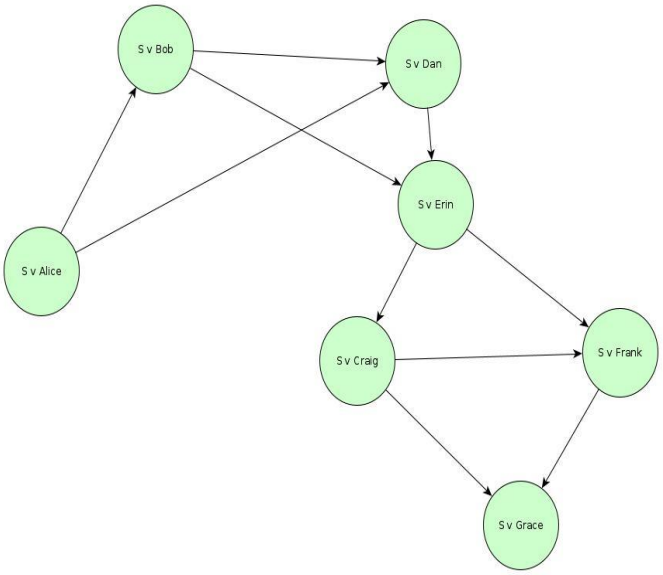

Figure 6: Citing-cases in a visual network graph.

From this diagram, we can quickly see that there is an unbroken line between *S v Alice* and *S v Grace.*

In order to determine whether there is a line of decisions between *S v Alice* and *S v Grace* using the tabular representation, the researcher must search as follows: first, she must check whether there is a direct citation. If there is not, then she must look through each of the tables for the cases cited by *S v Alice*, and check each of these. If *S v Grace* is still not found, then she should look through the tables of each of the cases cited by each of the cases cited by *S v Alice*, and so on.

It is easy to see that, in general, if there are n edges pointing outward from each node and the shortest path between two nodes being checked is m levels deep, then the worst-case complexity of the search will be n^m. For example, if each decision in a network has been cited exactly 10 times, and the second decision is 5 levels away from the first, then in the worst case the researcher may have to check as many as 100 000 rows in the tables to make the determination

5. Exploratory Data Analysis

At the outset I remarked that the network interface should be seen as an interface to support the use of citators for performing initial legal research in particular. The relationship between visualisation of data and initial research has also been recognised in the statistical sciences, most notably by John Tukey. Statistical research prior to Tukey likewise emphasised 'validation' of existing hypothesis. It is this view that underlies the practice of 'hypothesis testing', which will be familiar to many researchers. John Tukey emphasised a distinction between what he called "exploratory data analysis" and "confirmatory data analysis". In the latter endeavor, analytic methods are the primary tool of the statistician. However, when performing exploratory data

analysis, Tukey stressed the importance of data visualisation.

The underlying principle at play can be expressed in layman's terms as "needing to see the bigger picture". A view of data that focuses on isolated clusters of data cannot reveal broader trends and patterns at play. It would be akin to trying to study sociology by analysing the atoms that make up humans. It is simply not the correct level of analysis. It is the role of data visualisation to be able to present such a broad view of the data in a digestible format.

6. The Citator Application

I now present an example from the Citator[3] that demonstrates the broader objective of exploratory data analysis. These graphs are of course best viewed interactively on the website, but nonetheless some comments can be made based on figure 7 below.

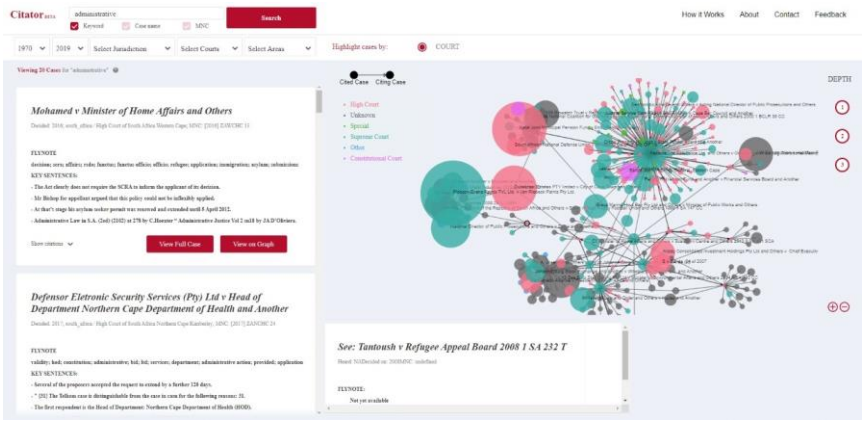

Figure 7. The Citator application

This graph shows the neighborhood of the case of *Mohamed v Minister of Home Affairs and Others*, with cases that are related to it by at most three degrees.

The Citator Beta allows the user to highlight nodes by court. The first full release of the Citator will also allow for highlighting by jurisdiction as well as by the dates of the decisions. We expect that highlighting by jurisdiction will be invaluable for comparative legal research, and indeed it is a distinguishing factor of the Citator that it has indexed cases from multiple jurisdictions in Africa.

The influence which a case has had is illustrated by the size of the node. At the time of writing, the metric being used to determine influence is simply the number of times that the case has been cited. It is common, however, in a network setting to want to consider not only the number of nodes connected to a given node, but also the influence of *those* nodes. A well known algorithm for including this information in the influence metric is the page rank algorithm. With court decisions in particular, however, users might also be especially

[3] The Citator is freely accessible at citator.africanlii.org

interested in including information about the superiority of the courts, which have cited a case, in the influence metric. In the first full release of the Citator we will include a number of different influence metrics and allow the user to select the one which is most appropriate for their use-case.

From figure 7 we are immediately able to glean insight into the jurisprudential landscape surrounding *Mohamed v Minister of Home Affairs* as a whole. Two notable aspects of the graph are that:

1. There appears to be a high proportion of decisions which have citations in common, but do not cite each other.
2. While there are clearly some very influential decisions, there are also many decisions which have been cited, but only a handful of times. This seems to indicate that so-called "leading decisions" may play less of a role as authority than legal researchers currently assume, or at least in this case.

Ultimately, the value of the graph is in providing the user with a broad view of the decisions on the topic and how they relate to each other. This may be of much help to a researcher who is quickly trying to get a grasp of the jurisprudential landscape on a point of law that they are not familiar with, and this is how we envision the citator being used.

7. Conclusion

In this paper I presented the role of citators as being twofold: for initial research, or exploratory data analysis, and for validation of research after the fact, or confirmatory data analysis. I argued that while tabular representations of citation networks may facilitate confirmatory analysis well, they are severely limited in the scope of information that they can display for the purposes of exploratory data analysis. In particular, relationships that result from the composition of citation relationships cannot feasibly be investigated using tabular representations. I showed how a network visualisation representation of citations is able to facilitate this research.

At the outset I made mention of the fact that the principle of *stare decisis* underlies the need for citators. The relationship, however, goes both ways. In *The Nature and Authority of Precedent* (Duxbury, 2008) Neil Duxbury argues that the rise of the principle of *stare decisis* was itself the product of judges increasingly writing down their decisions, and publishers developing more sophisticated tools for legal research. By building a tool that provides a unified citator interface to jurisprudence from across the African continent, I hope that we will be able to assist researchers in better understanding the development of precedent both within and between jurisdictions.

Related Works

The interested reader may find further information on network visualization by consulting the following:

1. Leibon, G. et al (2016) *Bending the Law*, available at: https://ssrn.com/abstract=2740136 (accessed 26 September 2019).
2. Lettieri, N. et al (2016) *A Computational Approach for the Experimental Study of EU Case Law: Analysis and Implementation*, Social Network Analysis and Mining, Vol. 5. No. 1 pp 1-17.
3. Tarissan, F. and Nollez-Goldbach, R. (2016) *Analysing the First Case of the International Criminal Court from a Network-Science Perspective*, Journal of Complex Networks, Vol 4, No.4, pp. 616–634

Appendix

Definition: Binary Relation

A binary relation is a triple (X, Y, R) in which $R \subseteq X \times Y$

Definition: Relation Composition

Let (X, Y, R_1) and (Y, Z, R_2) be relations. Then we define

$$R_2 \circ R_1 = \{(x, z) \in X \times Z \mid \exists y \in Y((x, y) \in R_1 \cap (y, z) \in R_2)\}$$

The quantifier in this definition makes clear the source of the computational complexity involved in determining composite relations.

Acknowledgements

I wish to thank my colleagues at the Democratic Governance and Rights Unit Mariya Badeva-Bright, Greg Kempe, and Amy Laura Sinclair, as well Danielle Louw for their helpful comments and suggestions on earlier drafts of this paper.

References

Dabney, L.C. (2007), *Citators: Past, Present, and Future*, Legal Reference Services Quarterly, Vol. 27, No. 2-3, pp. 165-190.
Duxbury, N. (2008), *The Nature and Authority of Precedent*, Cambridge.
ICLR (2019) *Judicial Consideration: a Reporter's Guide to Good Law*, available at: https://www.iclr.co.uk/knowledge/guides/judicial-

consideration-a-reporters-guide-to-good-law/ (accessed 26 September 2019).
Librarianship Studies, *Library and Information Science Encyclopedia*, available at: https://www.librarianshipstudies.com/2018/03/library-information-science-encyclopedia.html (accessed 31 July 2019).
Patti, J.O. (1993), *Mastering the Lawless Science of Our Law: A Story of Legal Citation Indexes*, Law Library Journal, Vol. 85, No. 1, pp. 1-48.
Washington University (2008) *Law Library Research Guides*, available at: https://libguides.law.wustl.edu/c.php?g=110786 (accessed 26 September 2019).